

## Aplikasi Penerjemah Bahasa Bangka Ke Bahasa Indonesia Menggunakan Neural Machine Translation Berbasis Website

Fadel Razsiah<sup>1\*</sup>, Ahmat Josi<sup>1</sup>, Sari Mubaroh<sup>1</sup>

<sup>1</sup> Politeknik Manufaktur Negeri Bangka Belitung, Sungailiat

\*E-mail : fadelrzh@gmail.com

Received: 10 Januari 2023; Received in revised form: 24 Januari 2023; Accepted: 31 Januari 2023

### Abstract

*At this time, machine translation is one of the options for translating languages, especially for people who want to learn a language. Bangka language is a language that is often used by the Bangka people for everyday life. There have been many developments for machine translation, but no one has yet developed a translation machine for the Bangka language. The development of a machine translation using Neural Machine Translation (NMT) and its implementation using the Flask microframework is the first step for the development of a machine translation for the Bangka language. This study aims to make and find out the translation results of the Bangka language translator application into Indonesian using the RNN translator model. The development starts from creating a parallel corpus of Bangka language to Indonesian and machine translation architecture with the RNN model. The research method starts from dataset collection, data preprocessing, modeling and training, system evaluation and implementation. From the results of the BLEU Scores evaluation, a value of 55.3% was obtained for Bangka to Indonesian. Furthermore, the model is implemented in the form of a website by utilizing the Flask framework, so that users can easily translate.*

**Keywords :** BLEU scores; Flaskmr; Corpus paralel; Machine translation; RNN.

### Abstrak

Pada saat ini mesin penerjemah merupakan salah satu pilihan untuk melakukan terjemahan bahasa, terutama bagi orang-orang yang ingin mempelajari suatu bahasa. Bahasa Bangka merupakan Bahasa yang sering digunakan masyarakat Bangka untuk keseharian. Telah banyak pengembangan mesin penerjemah, tetapi belum ada yang mengembangkan mesin penerjemah untuk bahasa Bangka. Pengembangan mesin penerjemah dengan *Neural Machine Translation* (NMT) dan implementasinya dengan menggunakan microframework Flask merupakan langkah awal untuk pengembangan mesin penerjemah bahasa Bangka. Penelitian ini bertujuan untuk membuat dan mengetahui hasil terjemahan aplikasi penerjemah bahasa Bangka ke bahasa Indonesia dengan model penerjemah RNN. Pengembangan dimulai dari membuat korpus paralel bahasa Bangka ke bahasa Indonesia dan arsitektur mesin penerjemah dengan model RNN. Metode penelitian dimulai dari pengumpulan *dataset*, *preprocessing* data, pemodelan dan pelatihan, evaluasi dan implementasi sistem. Dari hasil evaluasi BLEU Scores didapatkan nilai sebesar 55,3% untuk bahasa Bangka ke bahasa Indonesia. Selanjutnya model diimplementasikan ke bentuk website dengan memanfaatkan *framework* Flask, sehingga pengguna dapat dengan mudah dalam melakukan penerjemahan.

**Kata kunci:** BLEU Score; Flaskmr; Korpus paralel; Mesin penerjemah; RNN.

### 1. PENDAHULUAN

Penggunaan mesin penerjemah bahasa sangat sering kita gunakan dalam keseharian, selain untuk pembelajaran bahasa mesin penerjemah juga berperan

penting dalam pariwisata maupun perdagangan. Mesin penerjemah seperti, google translate juga mampu membantu para pelajar untuk belajar bahasa asing dengan lebih efisien. Namun, dengan

banyaknya bahasa daerah yang terdapat di Indonesia tidak mampu untuk dikelola oleh Google Translate itu sendiri termasuk bahasa Bangka. Indonesia mempunyai keragaman bahasa sebanyak 726 bahasa daerah dan 719 diantaranya masih sering digunakan dalam keseharian masyarakat Indonesia [1].

Mesin penerjemah merupakan pilihan utama setiap orang untuk sekedar mengetahui ataupun mempelajari suatu bahasa di luar bahasa keseharian mereka. Bahasa Bangka merupakan bahasa yang digunakan dalam aktivitas seluruh masyarakat Kepulauan Bangka [2]. Terkait Bahasa keseharian yang digunakan masyarakat Bangka adalah bahasa Bangka Melayu Bangka [3]. Bahasa Bangka memiliki beberapa dialek yaitu, dialek kota pelabuhan Mentok, dialek Kota Pelabuhan Belinyu, dialek Kota Pemerintahan Sungailiat dan dialek daerah pedesaan [4]. Pada penelitian ini menggunakan dialek Kota Pemerintahan Sungailiat.

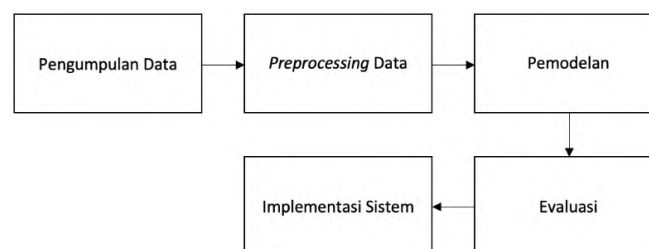
Terdapat beberapa penelitian dalam pengembangan model mesin penerjemah, seperti mesin penerjemah dengan *neural network* [5]. Mesin penerjemah dengan *neural network* dan mesin penerjemah dengan statis, dengan hasil penelitian bahwa model *recurrent neural network* (RNN) memiliki akurasi dan performa yang lebih baik dibandingkan dengan mesin penerjemah dengan statis [7]. Penelitian

lainnya juga mengatakan bahwa mesin penerjemah dengan model RNN memiliki performa yang lebih baik dari model penerjemah dengan model statis, ketika menggunakan sumber bahasa yang sedikit (*low resource language*) [8]. Hal ini membuktikan bahwa mesin penerjemah dengan model RNN dapat digunakan dalam mesin penerjemah Bahasa Bangka yang dapat dikategorikan memiliki sumber data bahasa yang sedikit (*low resource language*).

Penelitian ini bertujuan untuk merancang model mesin penerjemah berbasis RNN yang diimplementasikan kedalam bentuk website menggunakan *framework* Flask. Pada saat ini masih sedikit penelitian yang mengembangkan model mesin penerjemah dari bahasa daerah, khususnya bahasa Bangka ke bahasa Indonesia. Penelitian ini juga menghasilkan *dataset* korpus paralel dari Bangka ke bahasa Indonesia, yang telah digunakan dalam pengembangan mesin penerjemah bahasa Bangka ke bahasa Indonesia.

## 2. METODE PENELITIAN

Pada penelitian ini terdapat beberapa tahapan yang terdiri dari pengumpulan data, *preprocessing* data, pemodelan dan pelatihan, evaluasi, pengujian dan implementasi sistem. Gambar 1. mengilustrasikan dari alur pada penelitian ini.



Gambar 1. Flowchart Metode Penelitian

### 2.1 Pengumpulan Data

Pada penelitian ini menggunakan dataset berupa korpus paralel bahasa Bangka ke bahasa Indonesia yang bersumber dari buku yang berjudul "Kamus Bahasa Melayu Bangka - Indonesia" yang diterbitkan oleh Kementerian Pendidikan dan Kebudayaan pada tahun 2018. Data yang telah dikumpulkan disimpan dengan format

.csv untuk memudahkan pada proses *import* data.

### 2.2 Preprocessing Data

*Preprocessing* Data merupakan tahapan sebelum proses pengklasifikasian yang diperlukan untuk membersihkan, menghilangkan, mengubah sumber data, baik itu berupa karakter non alfabet maupun kata-kata yang tidak diperlukan [9]. Pada tahapan *preprocessing* pada *dataset*

bertujuan untuk menstrukturkan data yang belum terstruktur. Terdapat beberapa tahap untuk melakukan proses *preprocessing* pada dataset yang akan digunakan untuk memproses jalannya algoritma. Sebelum digunakan dataset harus melalui beberapa tahapan yaitu, *case folding*, *removing punctuation*, *removing digits* dan *tokenization*. Selain itu proses ini juga bertujuan untuk membersihkan data dari berbagai komponen atau fitur yang tidak diinginkan sehingga memaksimalkan pembelajaran mesin. *Preprocessing* data dilakukan untuk mempersiapkan data untuk digunakan selama proses pembelajaran mesin.

### 2.3 Pemodelan

Pemodelan dilakukan dengan memilih arsitektur dan varian model yang akan digunakan. Pada penelitian ini model yang telah dibuat akan dievaluasi. Model yang akan digunakan yaitu model RNN. Sedangkan untuk komponen pemrograman pada penelitian ini menggunakan bahasa pemrograman *python* dan *library tensorflow*. Pada model akan menggunakan *Hyperparameter* dengan jumlah *batch size* sebanyak 25, jenis *optimizer* yang digunakan adalah *Root-Mean-Square Propagation* (RMSprop). Untuk perhitungan *loss* menggunakan *sparse categorical entropy*. Selain itu penerapan *early stopping* juga digunakan dengan nilai *patience* sebesar 10 agar model akan terus melakukan pelatihan sebesar 10 kali secara beruntuk dan tidak menghasilkan progres *loss* melewati batas nilai nol.

### 2.4 Evaluasi

Pada tahapan ini peneliti melakukan evaluasi dengan menganalisa aktifitas grafik *loss* dan evaluasi dengan BLEU Score dari hasil terjemahan dengan otomatis [10]. Nilai BLEU Score yang tinggi menunjukkan bahwa terjemahan yang dihasilkan hampir menyamakan terjemahan dari refrensi. Pada tahap evaluasi ini akan memebandingkan hasil terjemahan yang telah dihasilkan oleh mesin penerjemah dengan terjemahan yang terdapat pada refrensi.

### 2.5 Implementasi Sistem

Pada tahapan ini akan mengimplementasikan model ke dalam aplikasi berbasis *website* dengan menggunakan *microframework* Flask. Perancangan *user interface* juga diterapkan agar pengerjaan aplikasi lebih terarah.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Pengumpulan Data

Pada penelitian ini menggunakan data berupa korpus paralel bahasa Bangka ke bahasa Indonesia. Data yang didapatkan pada "Kamus Bahasa Melayu Bangka - Indonesia" berjumlah 3080 data. Selanjutnya data yang didapatkan akan disimpan dengan format *file .csv* dan dilanjutkan untuk tahap *preprocessing* data dengan tahapan *case folding*, *removing punctuation*, *removing digits* dan *tokenization*. Pada Tabel 1. memperlihatkan contoh data yang berhasil dikumpulkan.

Tabel 1. Hasil Pengumpulan Data

| Korpus Paralel | Bahasa Bangka                               | Bahasa Indonesia    |
|----------------|---------------------------------------------|---------------------|
|                | Ka nek minum ape?                           | Kamu mau minum apa? |
|                | Ku yang pertama                             | Aku yang pertama    |
|                | Yo kita cube!                               | Mari kita coba!     |
|                | Ka dak ape-ape?                             | Kamu tidak apa-apa? |
|                | Jangan berbicara!                           | Jangan ngomong!     |
| Total Dataset  | 3080                                        |                     |
| Sumber         | Kamus Bahasa Melayu Bangka - Indonesia [11] |                     |

### 3.2 Hasil *Proprocessing* Data

Tahapan *preprocessing* terhadap data yaitu proses untuk membuat data yang belum terstruktur menjadi lebih terstruktur. Melalui proses *preprocessing* data, data yang dihasilkan dari pengumpulan data akan

di standarkan agar dapat diproses dengan algoritma yang akan digunakan pada penilitian. Sebelum data dapat diproses data harus dapat melalui beberapa tahapan seperti, *case folding*, *removing punctuation*, *removing digits* dan *tokenization*.

a. *Case Folding*

Hal ini dilakukan untuk membakukan data dengan kata yang sama tetapi kasus yang berbeda. Karena dataset berisi huruf besar dan huruf kecil, langkah ini dilakukan. Misalnya, frasa "Apa", "ApA", atau "aPa" dapat menjadi kesalahpahaman pada

komputer. Sehingga untuk menghindari hal tersebut pada penelitian ini menggunakan satu standar yaitu dengan menyamakan semua data dengan menjadikan data menjadi huruf kecil. Berikut Tabel 2. merupakan contoh dari hasil dari proses *case folding*.

Tabel 2. *Case Folding*

| Sebelum           |                     |
|-------------------|---------------------|
| Bahasa Bangka     | Bahasa Indonesia    |
| Ka nek minum ape? | Kamu mau minum apa? |
| Ku yang pertama   | Aku yang pertama    |
| Yo kita cube!     | Mari kita coba!     |
| Ka dak ape-ape?   | Kamu tidak apa-apa? |
| Jangan berbicara! | Jangan ngomong!     |
| Sesudah           |                     |
| Bahasa Bangka     | Bahasa Indonesia    |
| ka nek minum ape? | kamu mau minum apa? |
| ku yang pertama   | aku yang pertama    |
| yo kita cube!     | mari kita coba!     |
| ka dak ape-ape?   | kamu tidak apa-apa? |
| jangn berbicara!  | jangn ngomong!      |

b. *Removing Punctuation*

Tujuan dari *removing punctuation* atau penghapusan tanda baca adalah untuk menghilangkan karakter yang tidak akan mempengaruhi bagaimana terjemahan dilakukan. Akibatnya, proses penyaringan dilakukan pada tahap ini untuk mencari

tanda baca yang tidak berdampak nyata pada kalimat dan tanda baca yang dapat merusak makna dari sebuah kalimat jika ditempatkan secara tidak tepat. Tanda baca berikut disaring oleh peneliti dalam penelitian ini: ",&%? + -". Pada Tabel 3. merupakan contoh dari hasil proses *removing punctuation*.

Tabel 3. *Removing Punctuation*

| Sebelum           |                     |
|-------------------|---------------------|
| Bahasa Bangka     | Bahasa Indonesia    |
| ka nek minum ape? | kamu mau minum apa? |
| ku yang pertama   | aku yang pertama    |
| yo kita cube!     | mari kita coba!     |
| ka dak ape-ape?   | kamu tidak apa-apa? |
| jangn berbicara!  | jangn ngomong!      |
| Sesudah           |                     |
| Bahasa Bangka     | Bahasa Indonesia    |
| ka nek minum ape  | kamu mau minum apa  |
| ku yang pertama   | aku yang pertama    |
| yo kita cube      | mari kita coba      |
| ka dak apeape     | kamu tidak apaapa   |
| jangn berbicara   | jangn ngomong       |

c. *Removing Digits*

Proses *removing digits* atau menghapus angka memiliki tujuan hampir sama dengan proses menghapus

*punctuation*, namun proses menghilangkan angka lebih pada menghapus angka atau digit. Tabel 4 adalah perintah *code* untuk melakukan *removing digits*.

Tabel 4. *Removing Digits*

| Sebelum           |                     |
|-------------------|---------------------|
| Bahasa Bangka     | Bahasa Indonesia    |
| ka nek minum ape? | kamu mau minum apa? |
| ku yang pertama   | aku yang pertama    |
| yo kita cube!     | mari kita coba!     |
| ka dak ape-ape?   | kamu tidak apa-apa? |
| jangan berbicara! | jangan ngomong!     |
| Sesudah           |                     |
| Bahasa Bangka     | Bahasa Indonesia    |
| ka nek minum ape  | kamu mau minum apa  |
| ku yang pertama   | aku yang pertama    |
| yo kita cube      | mari kita coba      |
| ka nek minum ape  | kamu mau minum apa  |
| ku yang pertama   | aku yang pertama    |

d. *Tokenization*

Proses *tokenization* adalah operasi memisahkan teks menjadi potongan-potongan berupa token, bisa berupa potongan huruf, kata, atau kalimat,

sebelum dianalisis lebih lanjut. Dalam proses *tokenizer* menggunakan *word tokenizer* yaitu untuk memecah kalimat dalam bentuk kata per kata. Berikut hasil dari tahapan *tokenization* pada Tabel 5.

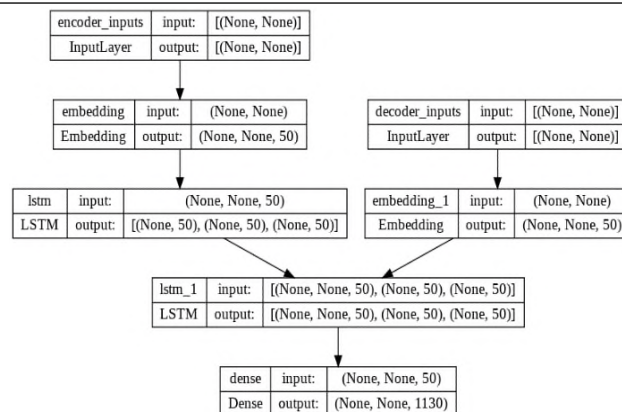
Tabel 5. *Tokenization*

| Sebelum                    |                             |
|----------------------------|-----------------------------|
| Bahasa Bangka              | Bahasa Indonesia            |
| jangan ngolokku            | jangan mengejekku           |
| dakde yang ngeliet kite    | tidak ada yang melihat kita |
| ku leteh bener             | aku sangat capek            |
| ape tu tembikar            | apa itu tembikar            |
| die orang yang tepat waktu | dia orang yang tepat waktu  |
| Sesudah                    |                             |
| Bahasa Bangka              | Bahasa Indonesia            |
| [24, 567]                  | [20, 632]                   |
| [195, 5, 33, 12]           | [10, 21, 6, 48, 11]         |
| [1, 95, 18]                | [4, 38, 132]                |
| [3, 6, 376]                | [3, 8, 811]                 |
| [10, 59, 5, 265, 73]       | [14, 71, 6, 148, 67]        |

3.3 Hasil Pemodelan

Pada penelitian ini model akan diprogram dari beberapa varian model agar model memiliki fungsi fleksibilitas. Program akan dijalankan dan program akan berfokus hanya pada satu modul yang dapat menampung pilihan pada setiap

*hyperparameter* dan dengan metode yang diinginkan, sehingga keseluruhan pada model akan otomatis disesuaikan. Penelitian ini memanfaatkan metode dari *subclassing* tensorflow dalam pengembangan program. Berikut hasil dari pemodelan dengan beberapa *hyperoaramater* pada Gambar 2.



Gambar 2. Arsitektur Model RNN

### 3.4 Hasil Evaluasi

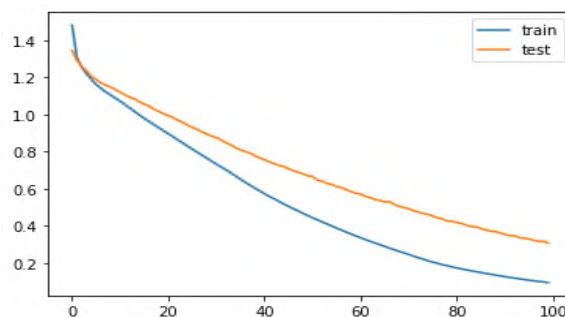
#### a. Aktifitas Grafik Loss

Pada saat proses training selesai, didapatkan nilai *loss* dan *validation loss* pada model penerjemah bahasa Bangka ke bahasa Indonesia yang telah dibuat. Nilai tersebut ditampilkan pada Tabel 6.

Berdasarkan tabel tersebut model memiliki nilai *loss* terendah 0.1 pada nilai *loss training* hingga 0.4 pada nilai *validation loss*. Semakin kecil nilai *loss* maka semakin baik kinerja model yang didapatkan. Berikut grafik nilai *loss* dan *validation loss* model penerjemah bahasa Bangka ke bahasa Indonesia Gambar 3.

Tabel 6. Aktifitas Grafik Loss

| Model Bangka - Indonesia |     |
|--------------------------|-----|
| Training Loss            | 0.4 |
| Validaton Loss           | 0.1 |



Gambar 3. Grafik Loss dan Validation Loss

#### b. BLEU Scores

Pada tahap evaluasi menggunakan BLEU Score bertujuan untuk menilai seberapa akurat hasil terjemahan yang dihasilkan oleh model mesin penerjemah dengan menggunakan BLEU Score. Hasil terjemahan akan dibandingkan dengan kalimat refrensi yang terdapat pada data korpus paralel. Hasil evaluasi membandingkan kalimat refrensi dibandingkan dengan kalimat hasil

terjemahan. Rentang penilaian BLEU Score adalah dari 0 - 1. Hasil terjemahan yang mendekati kalimat refrensi akan di nilai dengan skor 1, kalimat refrensi yang dimaksudkan adalah kalimat yang terdapat pada korpus paralel. Pada penelitian ini penentuan nilai BLEU Score menggunakan library NLTK (*Natural Language Toolkit Library*) dengan bahasa *python*. Berikut Tabel 7. menampilkan hasil dari evaluasi BLEU Score.

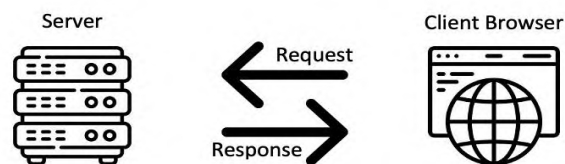
Tabel 7. Hasil Evaluasi BLEU Score

| Model Bangka - Indonesia |       |
|--------------------------|-------|
| Jumlah Dataset           | 3.080 |
| Epoch                    | 100   |
| Batch Size               | 28    |
| BLEU Scores              | 55,3% |

### 3.5 Hasil Implementasi Sistem

Pada tahap implementasi sistem peneliti menggunakan bahasa pemrograman *python* dengan menggunakan *microframework* Flask. Untuk arsitektur pada aplikasi dibagi menjadi dua yaitu, *client* dan *server*, *client* akan melakukan *request* dan menerima *response* dari *server*, pada hal ini *server* bertugas sebagai pemberi layanan pada aplikasi. Komunikasi yang terjadi antara

*client* dan *server* menggunakan metode REST API, sehingga memungkinkan *user* untuk melakukan *request* ke *server* melalui *url* yang telah disediakan oleh *server*, *server* kemudian mengirimkan *response* kepada *user* berdasarkan apa yang diminta oleh *user*. Pemilihan Flask menjadi *tools* dalam penelitian ini adalah karena Flask merupakan *microframework* berbasis *python* yang tepat untuk arsitektur REST API. Gambar 4. adalah ilustrasi arsitektur aplikasi yang digunakan dalam penelitian ini.



Gambar 4. Arsitektur Aplikasi

*Client browser* merupakan *browser* pada *client* yang berguna untuk *user* agar lebih mudah untuk melakukan *request* pada aplikasi. Sehingga pembuatan *user interface* diharapkan terlihat *simple* tetapi juga harus mementingkan fungsi dari aplikasi ini sendiri. Pada penelitian ini aplikasi dibuat

*single page* karena fungsi pada aplikasi ini hanya untuk menerjemahkan kalimat dari *input*-an yang dimasukkan *user*, sehingga aplikasi ini terlihat sederhana dan juga *user friendly*.

#### a. Pembuatan User Interface



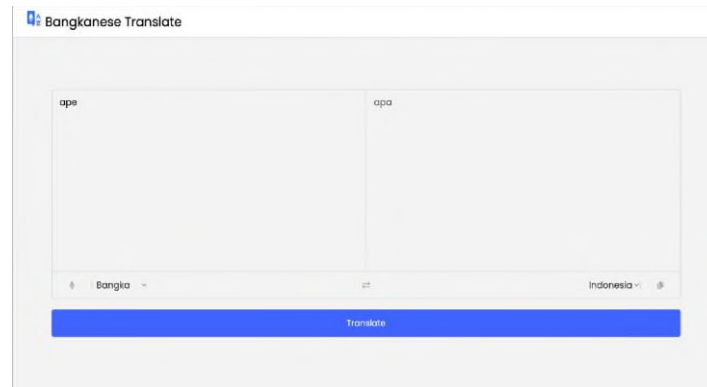
Gambar 5. User Interface

Pada Gambar 5. pada *user interface* menunjukkan bahwa aplikasi bersifat *one page* karena aplikasi hanya terdiri dari satu halaman. Pada halaman ini terbagi dua komponen penting yaitu pada *input* teks area dan *output* teks area, berikut penjelasan dari Gambar 5.

1. *Input text*, merupakan teks area yang digunakan *user* untuk memasukkan kalimat yang akan diterjemahkan.
2. *Output text*, merupakan teks area yang akan menampilkan hasil terjemahan dari kalimat yang telah dimasukkan *user* pada *input text*.
3. *Select language*, merupakan tempat *user* untuk memilih bahasa yang di

- input dan bahasa yang akan menjadi *output* atau hasil dari terjemahan.
4. *Button translate*, merupakan tombol untuk user akan menampilkan hasil dari terjemahan.

b. Tampilan Aplikasi



Gambar 6. Tampilan Aplikasi

Ketika *user* meng-input kalimat ke input teks area dan memilih bahasa sumber dan bahasa target, maka *user* dapat menekan tombol *translate* untuk melakukan proses penerjemahan. Saat proses penerjemahan sedang berlangsung maka akan menampilkan proses tampilan *loading* pada aplikasi yang menandakan bahwa aplikasi sedang berjalan. Setelah proses selesai maka akan muncul hasil terjemahan pada *output* teks area. Gambar 6. adalah ilustrasi hasil dari output terjemahan.

#### 4. SIMPULAN

Pengembangan Mesin Terjemah telah banyak di kembangkan oleh industri dan peneliti, namun masih sedikit untuk mengembangkan bahasa daerah khususnya bahasa Bangka. Pada penelitian ini menghasilkan model penerjemah dengan nilai BLEU Scores sebesar 55,32% pada model bahasa Bangka ke bahasa Indonesia, namun nilai BLEU Score pada model ini masih sangat jauh jika dibandingkan dengan Mesin Penerjemah lainnya. Diharapkan penelitian kedepannya dapat mencoba dengan model arsitektur lainnya seperti arsitektur LSTM dengan *Attention Mechanism* (mekanisme perhatian) [5] dan arsitektur BERT [12] pada penelitian ini juga menghasilkan korpus paralel bahasa Bangka ke bahasa Indonesia dengan jumlah data sebanyak 3080.

#### DAFTAR PUSTAKA

- [1] A. Ginting and N. AZ, "Penerjemah Dua Arah Bahasa Indinonesia Ke Bahasa Daerah," *Jurnal Telematika MKOM*, 2012.
- [2] P. Saputra and M. Afifulloh, "Pemetaan Penggunaan Bahasa Melayu Bangka", *Jurnal Ilmiah Bahasa dan Sastra Terakreditasi*, vol. 3, 2020.
- [3] S. D. Astuti, "Tafsir Budaya Asak Budu Ge Saro pada Masyarakat Bangka Belitung dalam Perspektif Pendidikan Anak," *Tawshiyah: Jurnal Sosial Keagamaan dan Pendidikan Islam*, vol. 15, pp. 86-101, 2020.
- [4] B. Nothofer, *Dialek Melayu Bangka, Malaysia: Universiti Kebangsaan Malaysia*, 1997.
- [5] D. Bahdanau, K. Cho and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *International Conference on Learning Representations*, 2015.
- [6] S. Yang, Y. Wang and X. Chu, "A Survey of Deep Learning Techniques for Neural Machine Translation," 18 February 2020. [Online]. Available: <https://arxiv.org/abs/2002.07526>.
- [7] R. Sennrich and B. Zhang, "Revisiting Low-Resource Neural Machine Translation: A Case Study," *ACL 2019 - 57th Annual Meeting of the Association*

- 
- for Computational Linguistics*, pp. 211-221, 2020.
- [8] F. A. Muttaqin and A. M. Bachtiar, "Implementasi Text Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak," *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, 2016.
- [9] K. Papineni, S. Roukos, T. Ward and W. J. Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation," 2002.
- [10] Kementrian Pendidikan dan Kebudayaan, *Kamus Bahasa Melayu Bangka - Indonesia*, Kantor Bahasa Kepulauan Bangka Belitung, 2018.
- [11] J. Zhu, Y. Xia, L. Wu, D. He, T. Qin, W. Zhou, H. Li and T.-Y. Liu, "Incorporating BERT into Neural Machine Translation," 17 February 2020. [Online]. Available: <https://arxiv.org/abs/2002.06823v1>.